

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-160289

(43)Date of publication of application : 23.06.1995

(51)Int.Cl.

G10L 3/00

G06F 3/16

(21)Application number : 05-305178

(71)Applicant : CANON INC

(22)Date of filing : 06.12.1993

(72)Inventor : GOMI TOMOHIRO

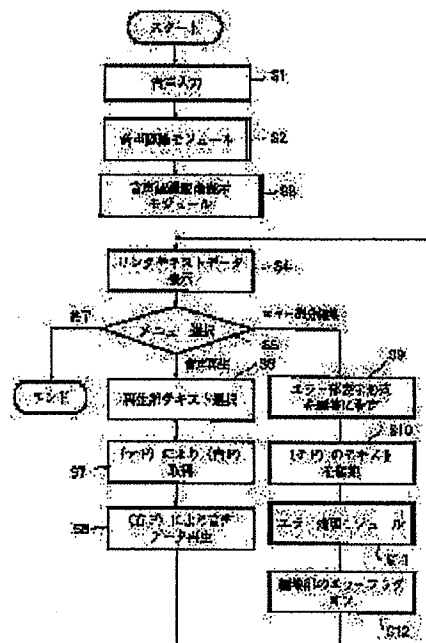
## (54) VOICE RECOGNITION METHOD AND DEVICE

## (57)Abstract:

**PURPOSE:** To provide a voice recognition method and a device which easily corrects a result recognized by a voice in correspondence to an inputted voice.

**CONSTITUTION:** An inputted voice is recognized and coded (S2), the recognized result is converted to characters based on the recognized result and displayed (S3). And a part which can not be recognized or a part which can not be specified is displayed with a phonetic symbol corresponding to a voice of the part (S11).

Thereby, the accurate recognized result of the part can be discriminated, and edition such as modification or correction of the part and the like can be performed.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-160289

(43) 公開日 平成7年(1995)6月23日

(51) Int.Cl. <sup>6</sup>	識別記号	片内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 6 1 C	9379-5H		
		D 9379-5H		
G 0 6 F 3/16	3 2 0 D	7323-5B		
		H 7323-5B		

審査請求 未請求 請求項の数6 OL (全 8 頁)

(21) 出願番号 特願平5-305178

(22) 出願日 平成5年(1993)12月6日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 五味 知宏

東京都大田区下丸子3丁目30番2号 キヤ  
ノン株式会社内

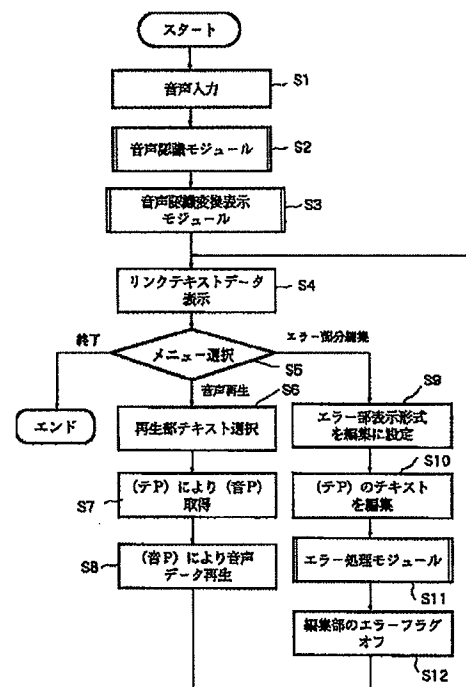
(74) 代理人 弁理士 大塚 康德 (外1名)

(54) 【発明の名称】 音声認識方法及び装置

(57) 【要約】

【目的】 音声認識した結果を、入力された音声に対応付けて容易に訂正できる音声認識方法及び装置を提供することを目的とする。

【構成】 入力された音声を認識してコード化し (S2)、その認識された結果に基づいて認識結果を文字に変換して表示する (S3)。そして、この表示された認識結果の内、認識できなかった箇所或いは特定できなかった箇所を、その部分の音声に対応した表音文字で表示する (S11)。これにより、その部分の正確な認識結果を判別することができ、その箇所の修正或いは訂正等の編集を行うことができる。



**【特許請求の範囲】**

**【請求項 1】** 音声を入力して認識する音声認識装置であって、

入力された音声認識してコード化する認識手段と、前記認識手段により認識された結果に基づいて認識結果を表示する表示手段と、前記認識結果の内、認識結果を特定できない箇所に対応した表音文字を表示する表音文字表示手段と、を有することを特徴とする音声認識装置。

**【請求項 2】** 前記表音文字表示手段に表示する文字の表記を指定する指定手段を更に有することを特徴とする請求項 1 に記載の音声認識装置。

**【請求項 3】** 前記入力された音声を録音する録音手段と、前記録音手段に録音された音声と前記認識結果を特定できない箇所とを対応付ける対応付手段と、前記特定できない箇所に対応する音声を前記録音手段より再生する再生手段とを更に有することを特徴とする請求項 1 に記載の音声認識装置。

**【請求項 4】** 前記表音文字表記手段は前記特定できない箇所を、元の音声の発音記号で示すことを特徴とする請求項 1 に記載の音声認識装置。

**【請求項 5】** 入力された音声認識して、その認識結果を表示する音声認識方法であって、入力された音声認識してコード化する工程と、その認識された結果に基づいて認識結果を表示する工程と、この表示された認識結果において、認識結果を特定できない箇所を指示する工程と、この指示された箇所に対応して表示する表音文字を指定する工程と、その指定された表音文字で前記箇所に対応する音声を表記して表示する工程と、を有することを特徴とする音声認識方法。

**【請求項 6】** 入力された音声認識して、その認識結果を表示する音声認識方法であって、入力された音声認識してコード化する工程と、その認識された結果に基づいて認識結果を表示する工程と、この表示された認識結果において、認識結果を特定できない箇所を指示する工程と、この指示された箇所に対応する音声を再生する工程と、を有することを特徴とする音声認識方法。

**【発明の詳細な説明】****【0001】**

**【産業上の利用分野】** 本発明は音声認識方法及び装置に関し、特に認識した結果を表示して修正できる音声認識方法及び装置に関するものである。

**【0002】**

**【従来の技術】** 音声を入力し、これを認識して文字等に交換して表示し、或いはその認識された音声に基づいて

データや各種命令を入力して処理する音声認識装置が知られている。このような音声認識装置では、音声の認識率の向上にばかり捕らわれ、使用者に対して使い勝手の良いユーザ・インターフェースについては、さほど注意が払われていないのが現状である。

**【0003】**

**【発明が解決しようとする課題】** 例えば、音声認識した結果が表示されている時、その認識結果のエラー部分或いは認識不能であった部分等を訂正したい場合がある。しかし、そのエラー部分等に対応する音声や、どういふ音声であったか分からないため、それを修正或いは訂正しようとしても、例えば前後の文章より判断するしかなかった。また、仮に、その入力した音声を録音しておいても、その録音されている音声のどの部分がエラーが発生した認識部分に対応しているかが分からず、その修正には多くの手間を要することになる。

**【0004】** 本発明は上記従来例に鑑みてなされたもので、音声認識した結果を、入力された音声に対応付けて容易に訂正できる音声認識方法及び装置を提供することを目的とする。

**【0005】** また本発明は、入力した音声とその認識結果とを容易に対応付けることができる音声認識方法及び装置を提供することを目的とする。

**【0006】**

**【課題を解決するための手段】** 上記目的を達成するために本発明の音声認識装置は以下の様な構成を備える。即ち、音声を入力して認識する音声認識装置であって、入力された音声認識してコード化する認識手段と、前記認識手段により認識された結果に基づいて認識結果を表示する表示手段と、前記認識結果の内、認識結果を特定できない箇所に対応した表音文字を表示する表音文字表示手段とを有する。

**【0007】** 上記目的を達成するために本発明の音声認識方法は以下の様な工程を備える。即ち、入力された音声認識して、その認識結果を表示する音声認識方法であって、入力された音声認識してコード化する工程と、その認識された結果に基づいて認識結果を表示する工程と、この表示された認識結果において、認識結果を特定できない箇所を指示する工程と、この指示された箇所に対応する音声を再生する工程とを有する。

**【0008】**

**【作用】** 以上の構成において、入力された音声を認識してコード化し、その認識された結果に基づいて認識結果を表示する。そして、この表示された認識結果の内、認識結果を特定できない箇所に対応した表音文字を表示するように動作する。

**【0009】**

**【実施例】** 以下、添付図面を参照して本発明の好適な実施例を詳細に説明する。

**【0010】** 図 1 は、本実施例の音声認識変換表示シス

テムにおける制御の流れを簡潔に表わした図である。

【0011】101は実施例の音声認識変換表示システム、102は音声認識システムで、入力された音声104を認識して音声コードデータ103を作成している。この作成された音声コードデータ103は音声認識変換表示システム101に出力され、テキストデータ表示105として出力されたり、或いは音声106で再生される。

【0012】図2は、本実施例の音声認識変換表示システム101の概略構成を示すブロック図である。

【0013】201は音声認識部（図3参照）であり、録音テープ或いはマイクロフォン等を含む音声入力装置208より入力された音声信号を認識し、対応する音声コードデータに変換する。202はフォント変換部で、各種言語に応じた文字パターンを発生させるためのフォントを有しており、コード化された音声データに基づいてフォント変換を行って対応する文字や記号等のパターンを作成する。203は、例えばCRT等の音声認識変換表示部で、音声認識部201で認識されて作成された音声コードデータがフォント変換部202によって文字パターン等に変換され、これらフォント変換された言語形式の文字パターン等を表示している。204はファイル管理部で、コード化された音声データを含むテキストファイルを管理している。205はコマンド制御部で、端末の使用者が入力する様々なコマンドを受け取り、それを解析してそのコマンドに基づく各種制御信号を発行している。このコマンド制御部205は、キーボードやマウス等のポインティングデバイスを備えている。206は補助記憶装置で、作成されたテキストファイル等の各種データを記憶している。207はバッファメモリで、例えば補助記憶装置206に保存される前の音声コードデータや、テキストデータ等を記憶している。

【0014】208は、例えばマイク、アンプ、スピーカ、電話機等を備えた音声入出力装置である。209は、CPU211により実行される制御手順を示すプログラム等記憶するROMである。210はRAMで、CPU211による各種制御の実行時にワークエリアとして使用され、各種データを一時的に記憶している。211はROM209に記憶された制御プログラムの手順に従って、装置全体を制御するCPUである。

【0015】図3は、図2の音声認識部201の概略構成を示すブロック図である。

【0016】301はサンプリング回路で、音声入出力装置208より入力した音声信号をサンプリングして保持している。302はA/D変換回路で、サンプリング回路301にサンプリングされて保持された音声信号（アナログ信号）をデジタル信号に変換している。303は特徴抽出回路で、デジタル信号に変換された音声データに基づいて、その音声データの特徴を解析する。304は比較回路で、特徴抽出回路303で解析された音

声データの特徴と、参照パターンデータ306に記憶されている基準音声パターンとを比較することにより、その音声データを認識している。305はバッファメモリで、この音声認識のために必要な各種データの記憶エリアを提供している。

【0017】以上の構成における本実施例の動作を図4～図6のフローチャートを参照して以下に詳細を説明する。尚、このプログラムを実行する制御プログラムはROM209に記憶されている。なお、この実施例では、音声信号は音声入出力装置208に含まれる留守番電話機等より入力される場合で説明している。

【0018】まずステップS1で、音声入出力装置208を介して音声信号が入力される。この音声信号は音声認識部201に送られ、サンプリング回路301でサンプリングされて、A/D変換器302でデジタル信号に変換され、特徴抽出回路303で認識された後、音声コードデータに変換される（ステップS2）。こうして認識され、作成された音声データは、ファイル管理部204を介して補助記憶装置206に送られて保存される。次にステップS3に進み、音声認識変換表示部203において、その音声コードデータを、フォント変換部202により、設定された言語形式のフォントに変換する。このステップS3の処理の詳細が、図5のフローチャートに示されている。

【0019】次に図5のフローチャートを参照して、このステップS3における音声認識変換表示処理を詳細に説明する。

【0020】ステップS21において、変換される言語が、例えば日本語、英語というように設定され、ステップS22では、この設定された言語に従って、認識された音声コードデータが文字コードに変換されてテキスト形式に変換される。ここで変換されたテキストと、録音されている音声データにおいて、それぞれ単語単位で位置を取得して対応付けられる（S23～S25）。即ち、ステップS23ではその単語の音声データにおける位置（音P）を取得し、ステップS24では、テキストデータにおける位置（テP）を取得する。そしてステップS26で、これら2つの位置情報をリンクする。尚、これら位置情報（音P）（テP）は、RAM210のワークエリアに記憶されている。

【0021】次にステップS26に進み、ステップS22で音声コードを文字フォントに変換する際にエラーが起きていたかどうかを判断し、エラーが発生していた場合はステップS27に進み、エラー（ERROR）処理モジュール（図6のフローチャート）を実行する。そしてステップS28に進み、音声データとテキストが対応付けられたリンクテキストデータを保存する。

【0022】次に、エラー処理モジュール（S11、S27）における処理を図6のフローチャートを参照して説明する。

【0023】例えば図5のステップS26でエラー有り  
と判断されると、ステップS31でエラーフラグがチェ  
ックされ、ステップS32では、そのエラーが発生して  
いる部分の音声コードの音声データにおける位置（音  
P）とテキストデータにおける位置（テP）とが取得さ  
れる。尚、このエラーフラグは、例えばRAM210  
に、そのエラーが発生した音声コードに対応付けて記憶  
されているものとする。

【0024】次にステップS33、S34において、エ  
ラー表示形式を、使用者が例えば「ローマ字」による表  
音形式を選択する等して設定する。次にステップS35  
に進み、エラーが発生した部分の音声コードを再取得し  
た後、ステップS36でこれを再び、ステップS34で  
指示された表音形式を用いて変換する。そしてステップ  
S37に進み、音声データにおける位置（音P）とテキ  
ストにおける位置（テP）を再リンクさせる。こうして  
エラー処理が終了する。

【0025】次に再び図4のフローチャートに戻り、以  
上のようにしてエラー表示形式が指定され、音声データ  
と対応付けられたテキストは、リンク・テキストデータ  
として表示される（ステップS4）。これにより、音声  
入力された音声信号は、端末上でテキストとして確認で  
きるようになる。

【0026】次に、その入力して認識された音声を、テ  
キストファイルより再生された音声として確認したい場  
合を想定する。

【0027】図4のステップS5のメニュー選択におい  
て、音声信号の再生が選択されるとステップS6に進  
み、再生したいテキストの一部または全体を選択する。  
次にステップS7～S8において、テキスト上の位置  
（テP）より音声信号における位置（音P）を求め、そ  
の位置（音P）より音声信号を読出して、その選択され  
たテキストの基になった音声データを直接、音声で再生  
することができる。これによりステップS8において、  
音声が入出力装置208を介して出力される。

【0028】次に、エラー部分の表示形式を設定或いは  
変更または、音声再生によって確認したエラー部分を訂  
正する等のリンクテキストを編集する場合を説明する。

【0029】図4のステップS5で、エラー部分の編集  
が指示されるとステップS9に進み、エラー部分の表示  
形式を編集に設定し、ステップS10で、そのエラー部  
分のテキストにおける位置（テP）を確保する（ステッ  
プS10）。そしてステップS11に進み、再びエラー  
処理モジュール（図6のフローチャート）を呼び出し  
て、エラー部分が再変換された後、確認・訂正されたテ  
キストを作成する。そしてステップS12で、エラーフ  
ラグをオフにして処理を終わる。これにより、エラー部  
分がエラー表記されるだけでなく、そのエラーの確認及  
び訂正を行うことができる。

【0030】以下に前述の処理を具体的な例を用いて説

明する。

【0031】〔実行例〕：日本語の音声信号を入力し、  
エラー表示に「ローマ字」モードを設定した場合を説明  
する。

【0032】入力した音声信号：『明日、11時に会議  
がありますので、よろしくお願い致します。』ここで、  
『明日』という音声部分が認識できなかった場合を説明  
する。前述の入力された音声信号をテキストに変換し、  
そのテキスト全体を表示するように指示すると、以下の  
ように表示される（ステップS34）。

【0033】変換表示：「ASUTA、11時に会議が  
ありますので、よろしくお願い致します。」ここで、  
「ASUTA」部分（エラー部分）を範囲指定して、そ  
の部分に対応している音声を再生すると（ステップS3  
5）、その再生音は『あすた』となる。ここで、そのエ  
ラー部分に対応する単語『ASHITA：（明日）』を入  
力して編集すると（ステップS36）、テキストの対  
応する部分が「明日」と変換され（ステップS37）、  
エラーフラグがオフになってエラーが解消される。

【0034】〔他の実施例〕以下、図面を参照して本発  
明の第2実施例を詳細に説明する。図における番号・名  
称は前述の実施例と同じである。

【0035】ここでも、音声入出力装置208の留守番  
電話等によって、音声データが入力された場合を想定す  
る。図2の音声入出力装置208を介して音声が入力さ  
れる（ステップS1）。この音声は、音声認識部201  
に送られ、音声認識部201により音声を認識する処理  
を行い、音声コードデータを作成する（ステップS  
2）。こうしてコード化されて録音された音声データ  
は、ファイル管理部204を介して補助記憶装置206  
に保存される。

【0036】次に、図2の音声認識変換表示部203に  
おいて、音声コードデータがフォント変換部202を介  
して設定された言語形式に変化される（ステップS  
3）。このステップS3の処理を詳細に説明したフロ  
ーチャートが図5である。ここで、ステップS3の処理、  
音声認識変換表示モジュールの処理を詳細に説明する。

【0037】ステップS21において設定された言語設  
定に従って、音声コードデータはフォント変換されてテ  
キスト形式に変換される（ステップS22）。ここで変  
換されたテキストと、録音されている音声データは、そ  
れぞれ単語単位で位置を取得され、対応付けられる（ス  
テップS23～S25）。次にステップS22のコード  
・フォント変換の際にエラーが起きていた場合は、ステ  
ップS26のエラー処理モジュールによって判断されて  
エラー処理モジュール（ステップS27）を行い、音声  
データとテキストが対応付けられたリンクテキストデー  
タを保存する（S28）。ここで、エラー処理モジュ  
ールについて、図6のフローチャートに従い説明する。

【0038】図5のステップS26でエラーとされた音

声コードデータは、図6のステップS31～37によって、変換テキストとの位置を確保される。

【0039】ここではエラー表示形式を、使用者が「ひらがな」の表音形式を選択して設定した場合で説明する。この設定されたエラー表示形式によって、音声コードを再取得した後（ステップS35）、これを再変換した（ステップS36）後に、音声データとテキスト部を再リンクさせる（ステップS37）。

【0040】以上のように作成されたエラー表示形式を指定され、音声データと対応付けられたテキストは、リンクテキストデータとして表示される（ステップS4）。これにより、該音声入力された音声は、端末上でテキストとして確認することが可能となる。

【0041】次に、音声として確認したい場合を想定する。図4のステップS5において音声再生を選択して、テキストの一部または全体を選択し（S6）、ステップS7～S8によって、選択したテキストの基になった音声データを直接に再生することができる。この場合、再生された音声は音声出力装置208を介して出力される。

【0042】この場合のエラー処理を具体例を用いて説明する。

【0043】【実行例】：英語の音声でエラー表示に「ひらがな」モードを設定した場合で説明する。

【0044】入力した音声：The teacher gave us a talk on human relations. That's the talk!

ここで、“The teacher”が認識できなかった場合で説明すると、エラー発生後の変換表示は、次のようになる。「てーちゃ gave us a talk on human relations. That's the talk!」。ここで、ひらがなで表示されたエラー部分「てーちゃ」を範囲指定し、その部分の音声を再生する。これにより、その再生音は、『the teacher』となる。この後、その再生された部分に該当する単語「The teacher」を入力することにより、そのエラー部分を訂正することができる。

【0045】次に同様にして、音声を英語で入力し、エラー表示を「カタカナ」モードに設定した場合で説明する。いま、入力された音声を『I can't do everything. I am only human.』とし、その音声認識されて英語で表示される場合、例えば「I can't do everything. I am only ヒーマン」のように表示されると、最後の単語「human」が認識できなかったことが分かる。そこで、この認識エラーが発生した箇所「ヒーマン」を指定して、その部分の再生を指示すると、入力した音声の『human』が発声される。そこで使用者は、この発声された音声に基づいて正しい単語を「human」を入力して、エラーが解消される。

【0046】また、前述の実施例以外にも、エラー表示形式を直接編集の表音形式として選択しても良い。

【0047】更に前記実施例の他にも、音声を表現する

のに相応しい表現、例えば「発音記号」等に設定して、表示することも可能である。また、それらの表音表示を選択して指定することにより、使用者が理解し易い表音表記に変更できる。

【0048】尚、本発明は複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用しても良い。また、本発明はシステム或は装置に、本発明を実施するプログラムを供給することによって達成される場合にも適用できることは言うまでもない。

【0049】以上説明したように本実施例によれば、音声を入力することにより、音声データをテキスト形式のデータに変換できる。

【0050】また、音声データを認識して文字形式で表示する際に、認識できなかった音声データを、オペレータが所望する表音形式で表すことができる。この表音表記としては、例えば「ローマ字」「発音記号」「ひらがな」「カタカナ」等が考えられ、これ以外にも所望の表音形式を自由に設定できる。

【0051】また、音声を認識して変換されたテキストファイルのうち、オペレータが確認したい部分を選択して、その部分を音声で再生させることにより、その音声認識結果が正しいかどうか確認することができる。また、その音声認識された結果を確認した後、その認識された結果に基づくコードデータを、必要に応じて訂正することができる。

【0052】

【発明の効果】以上説明したように本発明によれば、音声認識した結果を、入力された音声に対応付けて容易に訂正できる効果がある。

【0053】また本発明によれば、入力した音声とその認識結果とを容易に対応付けることができる効果がある。

【図面の簡単な説明】

【図1】本実施例の音声認識変換表示システムにおける制御の流れを簡潔に表わした図である。

【図2】本実施例の音声認識システムの概略構成を示すブロック図である。

【図3】本実施例の音声認識部の概略構成を示すブロック図である。

【図4】本実施例の音声認識変換表示システムにおける処理動作を示すフローチャートである。

【図5】図4のステップS3の音声認識変換表示モジュールにおける処理を示すフローチャートである。

【図6】図6及び図5の処理におけるエラー処理モジュールの処理を示すフローチャートである。

【符号の説明】

201 音声認識部

202 フォント変換部

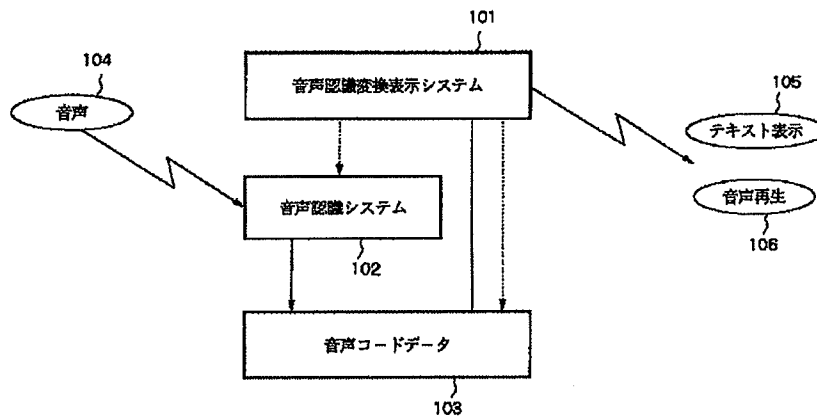
203 音声認識変換表示部

204 ファイル管理部

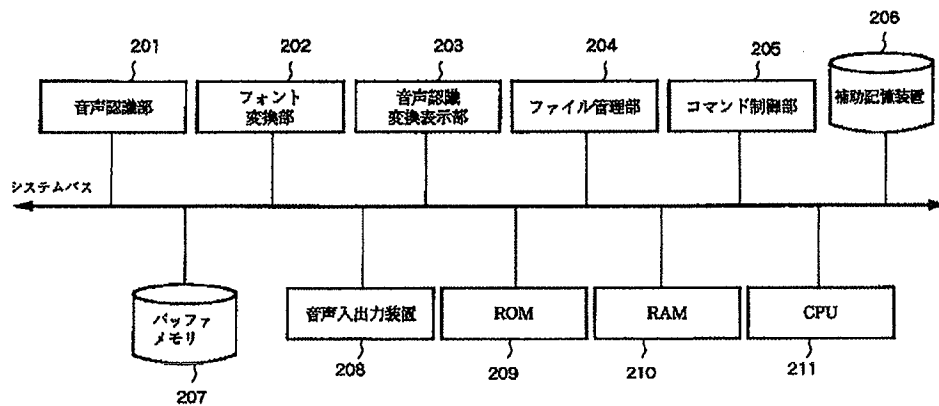
205 コマンド制御部  
206 補助記憶装置  
207 バッファメモリ  
208 音声入出力装置  
209 ROM  
210 RAM

211 CPU  
301 サンプル回路  
302 A/D変換回路  
303 特徴抽出回路  
304 比較回路  
306 参照パターンデータ

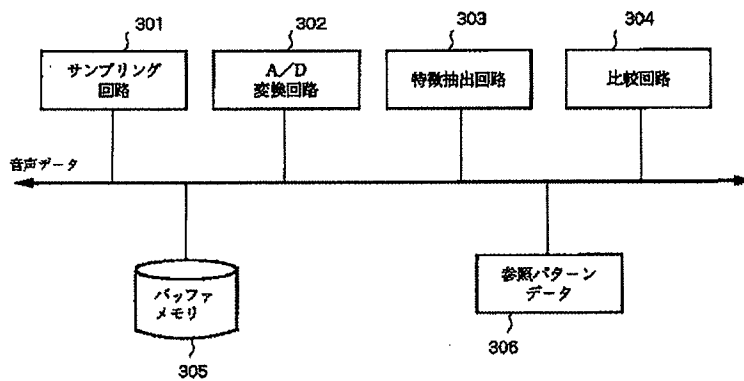
【図1】



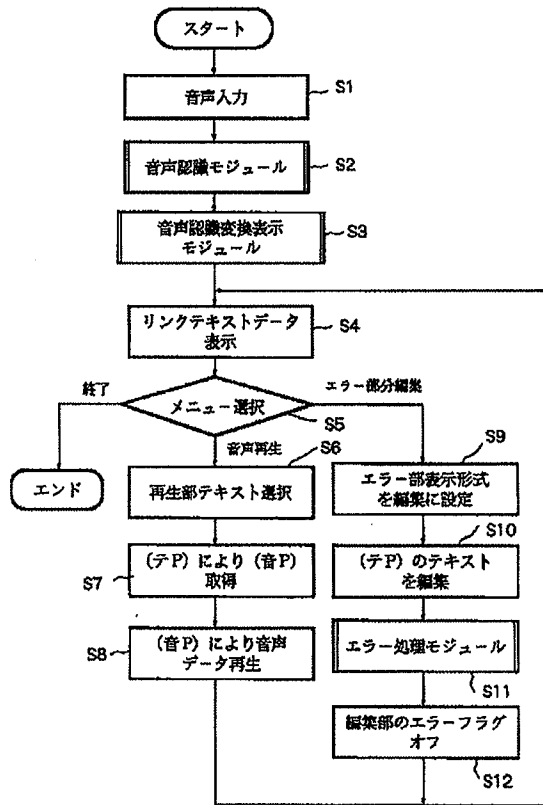
【図2】



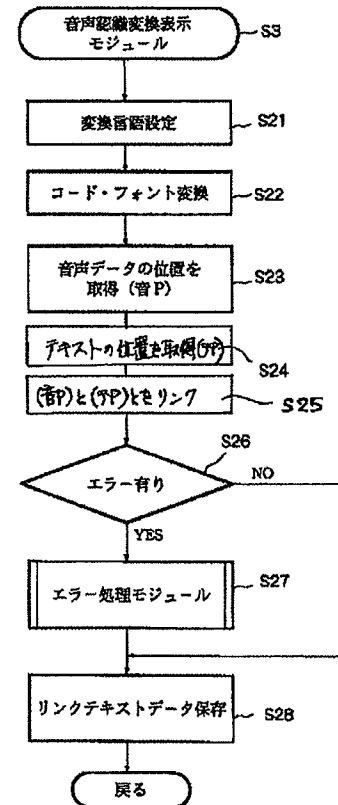
【図3】



【図 4】



【図 5】





【図 6】

